(54) Title: METHOD AND APPARATUS FOR INCREASING MEMORY RESOURCE UTILIZATION IN AN INFORMATION STREAM DECODER
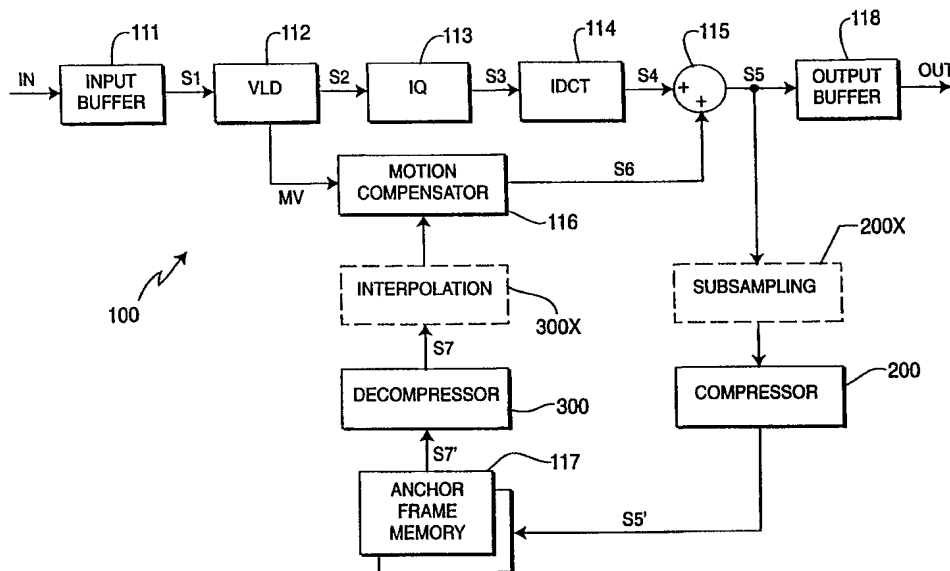
(57) Abstract

A method and apparatus for compressing one or more blocks of pixels using a Haar wavelet transform and a preferential quantization and scaling routine to produce one or more respective words comprising preferentially quantized Haar wavelet coefficients and associated scaling indicia. Within the context of an MPEG–like processing system, memory resource requirements, such as for frame storage, are reduced by a factor of two. In another embodiment, each pixel block is subjected to discrete cosine transform (DCT) processing and high order DCT coefficient truncation, thereby effecting an improved compression ratio of the pixel information.

METHOD AND APPARATUS FOR INCREASING MEMORY RESOURCE UTILIZATION IN AN INFORMATION STREAM DECODER

This application claims the benefit of U.S. Provisional Application

5   No. 60/084,632, filed 07-May-1998.

The invention relates to communications systems generally and, more particularly, the invention relates to a method and apparatus for increasing memory utilization in an information stream decoder, such as an MPEG-like video decoder.

10

BACKGROUND OF THE DISCLOSURE

In several communications systems the data to be transmitted is compressed so that the available bandwidth is used more efficiently. For example, the Moving Pictures Experts Group (MPEG) has promulgated several

15   standards relating to digital data delivery systems. The first, known as MPEG-1 refers to ISO/IEC standards 11172 and is incorporated herein by reference. The second, known as MPEG-2, refers to ISO/IEC standards 13818 and is incorporated herein by reference. A compressed digital video system is described in the Advanced Television Systems Committee (ATSC) digital television

20   standard document A/53, and is incorporated herein by reference.

The above-referenced standards describe data processing and manipulation techniques that are well suited to the compression and delivery of video, audio and other information using fixed or variable length digital communications systems. In particular, the above-referenced standards, and

25   other "MPEG-like" standards and techniques, compress, illustratively, video information using intra-frame coding techniques (such as run-length coding, Huffman coding and the like) and inter-frame coding techniques (such as forward and backward predictive coding, motion compensation and the like). Specifically, in the case of video processing systems, MPEG and MPEG-like

30   video processing systems are characterized by prediction-based compression encoding of video frames with or without intra- and/or inter-frame motion compensation encoding.

In a typical MPEG decoder, predictive coded pixel blocks (i.e., blocks that comprise one or more motion vectors and a residual error component) are decoded with respect to a reference frame (i.e., an anchor frame). The anchor frame is stored in an anchor frame memory within the decoder, typically a dual

5   frame memory. As each block of an anchor frame is decoded, the decoded block is coupled to a first portion of the dual frame memory. When an entire anchor frame has been decoded, the decoded blocks stored in the first portion of the dual frame memory are coupled to a second portion of the dual frame memory. Thus, the second portion of the dual frame memory is used to store the most recent full

10   anchor frame, which is in turn used by a motion compensation portion of the decoder as the reference frame for decoding predictive coded blocks.

Unfortunately, the cost of the memory necessary to implement an anchor frame memory may be quite high, in terms of money and in terms of integrated circuit die size (which impacts circuit complexity, reliability, power usage and

15   heat dissipation). Moreover, as the resolution of a decoded image increases, the size of the required memory increases accordingly. Thus, in a high definition television (HDTV) decoder, the amount of memory must be sufficient to implement the (dual) anchor frame anchor memory required to decode the highest resolution image. In addition, since the bandwidth requirements (i.e.,

20   speed of data storage and retrieval) of the frame store memory increase as the resolution of the image being decodes increases, the bandwidth of the frame store memory must be able to handle the requirements of the highest resolution image to be decoded.

Therefore, it is seen to be desirable to provide a method and apparatus

25   that significantly reduces the memory and memory bandwidth required to decode a video image while substantially retaining the quality of a resulting full-resolution or downsized video image.

SUMMARY OF THE INVENTION

30   The invention comprises a method and apparatus for reducing memory and memory bandwidth requirements in an MPEG-like decoder. Memory and memory bandwidth requirements are reduced by compressing image information

prior to storage, and decompressing the stored (compressed) image information prior to utilizing the image information in, e.g., a motion compensation module of the decoder. A method and apparatus are disclosed for compressing one or more blocks of pixels using a Haar wavelet transform and a preferential

5   quantization and scaling routine to produce one or more respective words comprising preferentially quantized Haar wavelet coefficients and associated scaling indicia. Within the context of an MPEG-like processing system, memory resource requirements, such as for frame storage, are reduced by a factor of two. In another embodiment, each pixel block is subjected to discrete cosine transform

10   (DCT) processing and high order DCT coefficient truncation, thereby effecting an improved compression ratio of the pixel information.

Specifically, one disclosed method according to the invention comprises the steps of: transforming, using a Haar wavelet transform, a pixel block to form a Haar coefficient block; quantizing, using a plurality of scaling factors, said

15   Haar coefficient block to form a respective quantized Haar coefficient block, said plurality of scaling factors being selected to preferentially allocate an available bit budget to quantized Haar coefficients representing relatively low vertical and horizontal spatial frequency components of said pixel block; and packing each of said quantized Haar coefficients and at least one of said plurality of scaling

20   factors into a respective word; and storing said word in a memory module within said decoding system.

BRIEF DESCRIPTION OF THE DRAWINGS

The teachings of the present invention can be readily understood by

25   considering the following detailed description in conjunction with the accompanying drawings, in which:

FIG. 1 depicts an embodiment of an MPEG-like decoder according to the invention;

FIG. 2 depicts a compression module suitable for use in the MPEG-like

30   decoder of FIG. 1;

FIG. 3 depicts a decompression module suitable for use in the MPEG-like decoder of FIG. 1;

FIG. 4 depicts a flow diagram of a data compression method according to the invention and suitable for use in the compression module of FIG. 1 or the MPEG-like decoder of FIG. 1;

FIG. 5 depicts a graphical representation of the data compression method of FIG. 4;

FIG. 6 depicts a block diagram of a compression process according to the invention; and

FIG. 7 depicts a block diagram of a compression process according to the invention.

To facilitate understanding, identical reference numerals have been used, where possible, to designate identical elements that are common to the figures.

DETAILED DESCRIPTION

The invention will be described within the context of a video decoder, illustratively an MPEG-2 video decoder. However, it will be apparent to those skilled in the art that the invention is applicable to any video processing system, including those systems adapted to DVB, MPEG-1 and other information streams.

Specifically, the invention will be primarily described within the context of an MPEG-like decoding system that receives and decodes a compressed video information stream IN to produce a video output stream OUT. The invention operates to reduce memory and memory bandwidth requirements in the MPEG-like decoder by compressing image information prior to storage, and decompressing the stored (compressed) image information prior to utilizing the image information in, e.g., a motion compensation module of the decoder.

FIG. 1 depicts an embodiment of an MPEG-like decoder 100 according to the invention. Specifically, the decoder 100 of FIG. 1 receives and decodes a compressed video information stream IN to produce a video output stream OUT. The video output stream OUT is suitable for coupling to, e.g., a display driver circuit within a presentation device (not shown).

The MPEG-like decoder 100 comprises an input buffer memory module 111, a variable length decoder (VLD) module 112, an inverse quantizer (IQ)

module 113, an inverse discrete cosine transform (IDCT) module 114, a summer

115, a motion compensation module 116, an output buffer module 118, an anchor

frame memory module 117, a compression module 200 and a decompression

module 300. Optionally, MPEG-like decoder 100 comprises a sub-sampling

5   module 200X and an interpolation module 300X.

The input buffer memory module 111 receives the compressed video

stream IN, illustratively a variable length encoded bitstream representing, e.g.,

a high definition television signal (HDTV) or standard definition television

signal (SDTV) output from a transport demultiplexer/decoder circuit (not

10   shown). The input buffer memory module 111 is used to temporarily store the

received compressed video stream IN until the variable length decoder module

112 is ready to accept the video data for processing. The VLD 112 has an input

coupled to a data output of the input buffer memory module 111 to retrieve, e.g.,

the stored variable length encoded video data as data stream S1.

15   The VLD 112decodes the retrieved data to produce a constant length bit

stream S2 comprising quantized prediction error DCT coefficients, and a motion

vector stream MV. The IQ module 113 performs an inverse quantization

operation upon constant length bit stream S2 to produce a bit stream S3

comprising quantized prediction error coefficients in standard form. The IDCT

20   module 114 performs an inverse discrete cosine transform operation upon bit

stream S3 to produce a bitstream S4 comprising pixel-by-pixel prediction errors.

The summer 115 adds the pixel-by-pixel prediction error stream S4 to a

motion compensated predicted pixel value stream S6 produced by the motion

compensation module 116. Thus, the output of summer 115 is, in the exemplary

25   embodiment, a video stream S5 comprising reconstructed pixel values. The video

stream S5 produced by summer 115 is coupled to the compression module 200

and the output buffer module 118. Optionally, the video stream S5 produced by

summer 115 is coupled to the sub-sampling module 200X.

The compression module 118 compresses the video stream S5 on a block

30   by block basis to produce a compressed video stream S5'. The operation of the

compression module 200 will be described in more detail below with respect to

FIGS. 2, 4 and 5. Briefly, the compression module 200 operates on a pixel block

by pixel block basis (e.g., a 4x4, 4x8 or 8x8 pixel block) to compress each pixel block within an anchor frame by processing the block according to a Haar wavelet transform, preferentially quantizing the resulting Haar coefficient block, scaling the resulting quantized coefficients and packing the scaled coefficients

5   and associated scaling factors to form a scaled, quantized, Haar domain representation of the pixel block that is coupled to the compression module output as part of compressed video stream S5'. It is noted by the inventors that such a scaled, quantized, Haar domain representation of the anchor frame requires approximately half the memory required for the pixel domain anchor

10 frame representation. Thus, the memory requirements of anchor frame memory module 117 are reduced by a factor of two.

Prior to performing a Haar wavelet transform on a pixel block, the pixel block may be processed using a one- or two-dimensional DCT function, and, optionally, a DCT coefficient truncation function. The resulting DCT coefficients

15 (truncated or not) are then processed according to the Haar wavelet transform.

In one embodiment of the invention, 4x4 pixel blocks are subjected to a 2D Haar wavelet transform, the results of which are adaptively quantized and scaled prior to being packed into data words for subsequent storage in, e.g., anchor frame memory module 117. Each of the resulting scaled, Haar domain

20 representations of the 4x4 pixel blocks occupies approximately 64 or fewer bits. This embodiment of the invention effects a 2:1 compression of pixel information, and will be discussed in more detail below with respect to FIGS. 4 and 5.

In another embodiment of the invention, 4x8 pixel blocks (i.e., 4 pixel rows by 8 columns) are processed using a 1D or 2D DCT to produce respective 4x8

25 blocks of DCT coefficients. After truncating the high order DCT coefficients, the resulting 4x4 blocks of DCT coefficients are subjected to 2D Haar wavelet transform, the results of which are adaptively quantized and scaled prior to being packed into data words. Each of the resulting scaled, Haar domain representations of the pixel blocks occupies approximately 64 or fewer bits. This

30 embodiment of the invention effects a 4:1 compression of pixel information, and will be discussed in more detail below with respect to FIG. 6.

The anchor frame memory module 117 receives and stores the compressed video stream S5'. Advantageously, the size of the anchor frame memory module 117 may be reduced to half the normal size, since a 2:1 compressed representation of the pixel block is stored, rather than the pixel block itself. The

5   compressed pixel blocks stored in the anchor frame memory module 117 must be provided to the motion compensation module 116 for further processing. However, since the motion compensation module processes uncompressed pixel data, the compressed pixel blocks are first provided to decompression module 300 via signal path S7' for decompression. Optionally, the output of the

10   decompression module 300 is coupled to interpolation module 300X for up-sampling (i.e., resizing) of the uncompressed pixel data prior to processing by motion compensation module 116.

Decompression module 300 essentially mirrors the operation of the compression module 200 described above. That is, decompression module 300

15   receives each Haar domain block and unpacks the received Haar domain block to retrieve the Haar coefficients and the previously used scaling factors. The unpacked Haar coefficients are then scaled using the unpacked scaling factors and subjected to an inverse Haar wavelet transform to substantially reproduce the original pixel block. Optionally, the inverse Haar wavelet transform

20   produces a block of DCT coefficients, which are then subjected to an inverse DCT transform to substantially reproduce the original pixel block. The decompression module 300 will be discussed in more detail below with respect to FIG. 3.

The decompression module 300 is accessed by the motion compensation

25   module 116 via signal path S7. That is, the motion compensation module 116 utilizes one or more stored anchor frames (e.g., the substantially reproduced pixel blocks forming the most recent I-frame or P-frame of the video signal produced at the output of the summer 115), and the motion vector signal MV received from the VLD 112, to calculate the values for the motion compensated

30   predicted pixel value stream S6.

The operation of optional sub-sampling module 200X and interpolation module 300X will now be descried. Sub-sampling module 200X operates to

operated to sub-sample, or decimate, the pixel information within video stream S5 to produce effect a resizing (i.e., downsizing) of the video image represented by the pixel data. For example, a 1920 pixel by 1080 line high definition image may be resized to a smaller image, such as a standard 720 pixel by 480 line

5   image, or some other non-standard format selected to generally reduce memory requirements or reduce memory requirement to a specific level determined by, e.g., available semiconductor area or other factors. The sub-sampled video stream is then processed by compression module 200 in the manner described elsewhere in this application.

10   Interpolation module 300X is used to resize (i.e., upsize) the previously downsized image information resulting from the operation of sub-sampling module 200X. The interpolation module 300X utilizes the image information provided by decompression module 300 to calculate (i.e., approximate) image information previously removed by the operation of the sub-sampling module

15   200X, such that the motion compensation module 116 receives pixel information representative of an appropriate image size. It must be noted that the operation of sub-sampling module 200X and interpolation module 300X will inherently cause a loss in image resolution and image quality. However, within the context of, e.g., a presentation system incapable of providing such resolution (i.e., a small

20   display tube television system) the loss of image resolution will not be noticed, and the cost savings from memory reduction may be critical to the market success of the presentation system.

       FIG. 2 depicts a compression module 200 suitable for use in the MPEG-like decoder of FIG. 1. The compression module 200 comprises an

25   optional discrete cosine transform module (DCT) 210, a wavelet transform module 220, illustratively a Haar wavelet transform module (HAAR), a quantization module (Q) 230, a packing module (PACK) 240 and a rate control module 250. The functionality of the compression module may be implemented as an application specific integrated circuit (ASIC) or as a general purpose

30   computer that is programmed to perform specific control functions in accordance with the present invention.

The compression module 200 operates in one of a discrete cosine transform (DCT) compression mode and a non-DCT compression mode to effect a compression of the pixel information included within the video stream S5.

In the non-DCT compression mode, the compression unit 200 effects a 2:1

5   compression of pixel information. For example, in the case of a 4x4 pixel block having 8 bit dynamic range (comprising approximately 128 bits of information), the compression module operating in the non-DCT mode produces a 64 bit representation of each 4x4 pixel without substantial loss in image quality. Similarly, in the case of an 8x8 pixel block (approximately 512 bits) or an 8x4

10   pixel block (approximately 256 bits), the compression module operating in the non-DCT mode produces, respectively, a 256 bit or a 128 bit representation without substantial loss in image quality.

In the DCT compression mode (i.e., where high order DCT coefficients are truncated prior to Haar transformation), the compression unit 200 effects a 4:1

15   compression of pixel information. Thus, in the case of an 8x8 pixel block (approximately 512 bits) or an 8x4 pixel block (approximately 256 bits), the compression module operating in the DCT mode produces, respectively, a 128 bit or a 64 bit representation of the pixel information, albeit with a slight reduction in high frequency image information (e.g., "edge" information).

20   The output of the DCT module 210 is coupled to the Haar transform module 220, where the input pixel block or DCT coefficient block is subjected to a Haar transform process. The resulting Haar coefficient block is coupled to the quantizer 230. Quantizer 230, in response to a control signal RC produced by rate controller 250, preferentially quantizes and scales the Haar wavelet

25   coefficient block to produce a data word, which is coupled to the packing module 240 as information stream S5'. Rate controller 250 monitors the input to the quantizer 230 and the output from the quantizer 230 to determine the type and size of the coefficient blocks being quantized.

In the non-DCT compression mode, each, illustratively, 4x4 pixel block

30   (i.e., pixels x_0 through x_{15}) is subjected to a two dimensional Haar wavelet transformation to produce a corresponding 4x4 Haar coefficient block (i.e., coefficients X_0 through X_{15}). Each of the Haar coefficients represents the spectral

energy associated with a portion of the horizontal spatial frequency components and vertical frequency components of the video information contained within the 4x4 pixel block. The Haar coefficients are then quantized in a preferential manner such that "important" coefficients (i.e., those coefficients representing

5   spatial frequency components providing image information more readily discernable to the human eye) are allocated more bits than less important coefficients. The quantized coefficients are then scaled by a factor of $1/2^n$, where n is selected as the minimum n that satisfies the following relationship: $2^{(n+m)}$ is less than the maximum value of the absolute value of wavelet coefficients in a

10 set, where a set refers those coefficients that share the same scaling factor and m is the number of bits allocated to each coefficient in the set. Finally, the scaled coefficients and the associated scaling factors are packed into an (approximately) 64 bit word. The 64 bit word is then coupled to the anchor frame memory module 117 as part of compressed video stream S5'.

15   In the DCT compression mode, each, illustratively, 4x8 pixel block (i.e., pixels x_0 through x_{31}) is subjected to a 1D or 2D DCT function to produce a corresponding 4x8 DCT coefficient block (i.e., coefficients X_0 through X_{31}). The high order DCT coefficients are the truncated (i.e., coefficients X_{16} through X_{31}), resulting in a 4x4 DCT coefficient block (i.e., coefficients X_0 through X_{15}). The

20 resulting DCT coefficient block is then subjected to a two dimensional Haar wavelet transformation to produce a corresponding 4x4 Haar wavelet coefficient block (i.e., coefficients X_0 through X_{15}). The Haar wavelet coefficients are then preferentially quantized, scaled and packed in substantially the same manner as provided in the non-DCT mode of operation to produce, in the case of a 4x8 pixel

25 block, a corresponding 64 bit word. The 64 bit word is then coupled to the anchor frame memory module 117 as part of compressed video stream S5'.
The selection of DCT mode or non-DCT mode may be made according to the needs of a particular application. For example, in a low resolution or standard definition (i.e., not high definition) television application, it may be acceptable to

30 "drop" some of the high frequency detail information in the image, whereas such information may be crucial to a high definition application. Additionally, the

10

memory constraints within which a decoder is designed may dictate the use of an enhanced level of compression, such as provided by the DCT mode of operation.

FIG. 3 depicts a decompression module 300 suitable for use in the MPEG-like decoder of FIG. 1. The decompression module 300 of FIG. 3

5   comprises a series coupling in the order named of an unpacking module 310, an inverse quantization (Q-1) module 320, an inverse Haar transform (HAAR-1) module 330 and an optional inverse discrete cosine transform module (IDCT) 340. The decompression module operates in the reverse order, and with the inverse functionality, of the corresponding modules described above with respect

10   to the compression module 200. Therefore, the operation of the decompression module will not be discussed in great detail. Briefly, the decompression module 300 receives a packed data word, unpacks the data word to retrieve a Haar wavelet coefficient block and performs an inverse Haar wavelet transformation on the unpacked coefficient block to produce a DCT coefficient block or a pixel

15   block. In the case of a DCT coefficient block that was not truncated by the compression module 200, the DCT coefficient block is subjected to an inverse DCT operation to produce a pixel block. In the case of a DCT coefficient block that was truncated by the compression module 200, NULL coefficients are added to the DCT coefficient block to fill it out to the appropriate size, and the "filled

20   out" DCT coefficient block is subjected to an inverse DCT operation to produce a pixel block.

The operation of the compression module 200 will now be described in more detail with respect to FIG. 4 and FIG. 5. FIG. 4 depicts a flow diagram of a data compression method according to the invention and suitable for use in the

25   compression module of FIG. 2 or the MPEG-like decoder of FIG. 1. FIG. 5 depicts a graphical representation of the data compression method of FIG. 4. The compression routine 400 of FIG. 4 may be implemented as a logical function between cooperating modules of the MPEG-like decoder 100 of FIG. 1. The compression routine 400 may also be implemented as a routine within a general

30   purpose computing device. The compression routine 400 operates to compress each, illustratively, 4x4 pixel block within a video stream (i.e., the video stream S5 produced at the output of adder 115 in the MPEG-like encoder of FIG. 1) to

produce a data set representative of the 4x4 pixel block with a much smaller amount of data.

The compression routine 400 is entered at step 402, when, e.g., a 4x4 pixel block (reference numeral 510 in FIG. 5) is received by compression module 200.

5   The compression routine 400 proceeds to step 403, where a query is made as to the mode of operation of the compression module. If the answer to the query as step 403 indicates that the compression module is operating in a non-DCT mode of operation, then the routine 400 proceeds to step 406. If the answer to the query as step 403 indicates that the compression module is operating in a DCT

10 mode of operation, then the routine 400 proceeds to step 404.

At step 404 a discrete cosine transform (DCT) is performed on the 4x4 pixel block to produce a 4x4 block of DCT coefficients (reference numeral 520 in FIG. 5). The DCT may be a one-dimensional or two-dimensional DCT. The compression routine 400 then proceeds, optionally, to step 405, where the high

15 order DCT coefficients generated by the DCT transform of step 404 are truncated (not shown in FIG. 5). The routine 400 then proceeds to step 406.

The combination of DCT transformation (step 404) and high order DCT coefficient truncation (step 405) is especially useful where greater compression is needed. For example, consider the case of input pixel blocks comprising 4x8

20 pixel blocks (i.e., 4 pixel rows by 8 columns) rather than 4x4 pixel blocks, where the memory allocation remains constant (i.e., a 4:1 compression is necessary). After performing, e.g., a 2D DCT as step 404 to produce a respective 4x8 block of DCT coefficients, the high order DCT coefficients (i.e., the 16 DCT coefficients representing the highest pixel spectra within the pixel block) are optionally

25 truncated to produce a 4x4 DCT coefficient block. This 4x4 DCT coefficient block comprises t to form a he majority of the pixel domain information contained within the pixel block, lacking only the highest frequency pixel domain information. To reconstruct the pixel domain information (e.g., at decompression module 300), 16 NULL DCT coefficients are added to the 16 non-truncated DCT

30 coefficients to form a 4x8 DCT coefficient block that, when processed according to an inverse DCT module, will produce a 4x8 pixel block. The produced 4x8 pixel block may lack some of the high frequency (i.e., detail) information found in the

original 4x8 pixel block, but such information may not be crucial to the application (or the memory to hold such information may not be available).

In the case of a 4x4 pixel block or an 8x8 pixel block, the high order truncation of step 405 will result in, respectively, a 2x4 DCT block or a 4x8 DCT

5   block. The number of coefficients truncated may be adapted such that more or less DCT coefficients are truncated. For example, the 8x8 DCT coefficient block resulting from a 2D DCT processing of an 8x8 pixel block may be truncated to produce a 4x4 pixel block (with a corresponding decrease in high frequency image information).

10   At step 406 a 2-dimension Haar wavelet transform is performed on the original 4x4 pixel block (510), the DCT coefficient block produced at step 404 (520), or, optionally, the truncated DCT coefficient block produced at step 405. The 2-dimensional Haar wavelet transform results in a 4x4 block of wavelet coefficients (reference numeral 530 in FIG. 5). Each of the wavelet coefficients

15   represents a specific frequency range of the vertical and horizontal spatial frequency components of the 4x4 pixel block 510.

Referring now to FIG. 5, the 4x4 block of wavelet coefficients (530) is divided into four quadrants according to the represented vertical and horizontal spectra. An upper left quadrant, denoted as an "LL" quadrant, includes those

20   wavelet coefficients representing low vertical spatial frequency components and low horizontal spatial frequency components of the 4x4 pixel block 510. An upper right quadrant, denoted as an "LH" quadrant, includes those wavelet coefficients representing low vertical spatial frequencies and high horizontal spatial frequencies. A lower left quadrant, denoted as an "HL" quadrant,

25   includes those wavelet coefficients representing high vertical spatial frequency components and low horizontal spatial frequency components. A lower right quadrant, denoted as an "HH" quadrant, includes those wavelet coefficients representing high vertical spatial frequency components and high horizontal spatial frequency components.

30   The LL quadrant is further divided into four sub-quadrants according to the represented vertical and horizontal spectra within the LL quadrant. An upper left sub-quadrant, denoted as an "LLLL" sub-quadrant, includes those

13

coefficients representing a substantially DC vertical and horizontal spatial frequency component of the 4x4 pixel block 510. An upper right sub-quadrant, denoted as an "LLLH" sub-quadrant, includes those wavelet coefficients representing low vertical spatial frequencies and high horizontal spatial

5   frequencies. A lower left sub-quadrant, denoted as an "LLHL" sub-quadrant, includes those wavelet coefficients representing high vertical spatial frequency components and low horizontal spatial frequency components. A lower right sub-quadrant, denoted as an "LLHH" sub-quadrant, includes those wavelet coefficients representing high vertical spatial frequency components and high

10 horizontal spatial frequency components.

After performing the 2D Haar wavelet transform (step 406), the compression routine 400 proceeds to step 408, where the wavelet coefficients are quantized using a 2^n scaling factor (reference numeral 540 in FIG. 5). The quantization process reduces the accuracy with which the wavelet coefficients

15 are represented by dividing the wavelet coefficients by a set of quantization values or scales with appropriate rounding to form integer values. The quantization values are set individually for each set of wavelet coefficients. By quantizing the DCT coefficients with this value, many of the DCT coefficients are converted to zeros, thereby improving image compression efficiency.

20 Specifically, the wavelet coefficients are preferentially quantized, such that "more important" coefficients (i.e., those wavelet coefficients representing spatial frequency components providing image information more readily discernable to the human eye) are allocated more bits than "less important" coefficients. That is, the number of bits used to represent each coefficient is reduced by an amount

25 inversely proportional to the relative importance of the coefficient in representing an image.

Since the human eye is more sensitive to low spatial frequency image components, the DC component of an image is the most important. Thus, the X_0 coefficient (i.e., the LLLL sub-quadrant) is quantized using a scaling factor of

30 four to an eight bit level of precision, resulting in a quantized DC component X_{0Q}. The remaining LL quadrant coefficients, i.e., the X_1, X_2 and X_3 coefficients are each quantized using a scaling factor of $4*2^m$ to a four bits (plus a sign bit) level

14

of precision, resulting in quantized LL components X_{1Q}, X_{2Q} and X_{3Q}. The scaling factor "m" comprises a 2-bit integer.

In the exemplary embodiment, the LH quadrant is deemed by the inventors to be the second most important quadrant. However, the quantization of the LH quadrant will depend upon how much of the 64 bit word budget is consumed in quantizing the remaining two quadrants (i.e., the HL and HH quadrants). Thus, the two "least important" quadrants, the HL and HH quadrants, are processed before the LH quadrant.

The HL quadrant coefficients, i.e., X_8, X_9, X_{12} and X_{13}, are each quantized using a scaling factor of $4*2^n$ to a two bit (plus a sign bit) level of precision, resulting in quantized HL components X_{8Q}, X_{9Q}, X_{12Q} and X_{13Q}. The HH quadrant coefficients, i.e., X_{10}, X_{11}, X_{14} and X_{15}, are each quantized using a scaling factor of $4*2^n$ to a two bit (plus a sign bit) level of precision, resulting in quantized HH components X_{10Q}, X_{11Q}, X_{14Q} and X_{15Q}. If quantization of a coefficient results in a zero value coefficient, then a sign bit is not used for that coefficient. The scaling factor "n" comprises a 3-bit integer.

After quantizing the coefficients for the LL, HL and HH quadrants (step 408), the compression routine 400 proceeds to step 410, where a query is made as to whether the number of zero-coded coefficients is greater than 3. That is, whether four or more of the coefficients quantized at step 408 were quantized to a value of zero, such that four or more sign bits were not used.

If the query at step 410 is answered affirmatively, then the compression routine 400 proceeds to step 414 where the LH quadrant coefficients, i.e., X_2, X_3, X_6 and X_7, are each quantized using a scaling factor of $4*2^{(n-1)}$ to a three bit (plus a sign bit) level of precision, resulting in quantized LH components X_{2Q}, X_{3Q}, X_{6Q} and X_{7Q}. The compression routine 400 then proceeds to step 416.

If the query at step 410 is answered negatively, then the compression routine 400 proceeds to step 412 where then the LH quadrant coefficients, i.e., X_2, X_3, X_6 and X_7, are each quantized using a scaling factor of $4*2^n$ to a two bit (plus a sign bit) level of precision, resulting in quantized LH components X_{2Q}, X_{3Q}, X_{6Q} and X_{7Q}. The compression routine 400 then proceeds to step 416.

At step 416, the quantized wavelet coefficients X_{0Q} through X_{15Q} are packed (reference numeral 550 in FIG. 5) to produce compressed output stream S5'. That is, in the case of the exemplary 4x4 pixel block processing described above, an approximately 8-byte (64 bit) word is formed by concatenating the scaled

5   coefficients and scaling factors as follows: "X_{0Q}, m, X_{1Q} X_{2Q}, X_{3Q}, X_{3Q}, n, X_{4Q}, X_{5Q}, X_{6Q}, X_{7Q}, X_{8Q}, X_{9Q}, X_{10Q}, X_{11Q}, X_{12Q}, X_{13Q}, X_{14Q}, X_{15Q}".

After packing the scaled coefficients and scaling factors (step 416), the routine 400 proceeds to step 418, where a query is made as to whether there are more pixel blocks within the frame to be processed. If the query at step 418 is

10   answered affirmatively, then the routine 400 proceeds to step 406 (or optional step 404), and the above-described transform, quantization scaling and packing process is repeated for the next pixel block. If the query at step 418 is answered negatively, then the routine 400 proceeds to step 420 and exits. The routine may be reentered at step 402 when, e.g., the first pixel block of a subsequent image

15   frame is received.

It should be noted that the order of packing is not relevant to the practice of the invention. However, since certain coefficients and scaling factors are required to be utilized before others for, e.g., inverse scaling and transform operations, the inventors have identified the disclosed order as a preferred order

20   within the context of an MPEG-like video information stream. Packing order may be differently optimized for other types of information streams.

In the exemplary embodiment of the invention a 4x4 pixel block is processed in a non-DCT mode, and a 4x8 pixel block is processed in a DCT mode of operation. Therefore, the above described sub-quadrants each contain only

25   one wavelet coefficient. It will be readily understood by those skilled in the art that since blocks of any size may be processed according to the invention, the sub-quadrants may comprise more than one wavelet coefficient. Moreover, the quadrants and sub-quadrants may comprise a non-symmetrical grouping of wavelet coefficients.

30   A very important case discussed above is the 4x8 pixel block case. As discussed above with respect to FIG. 4, the 4x8 pixel block may be subjected to DCT processing and truncation prior to Haar wavelet transformation to produce

a 4x4 Haar wavelet coefficient block. If the optional step of DCT coefficient
truncation is not used, then each quadrant of the Haar wavelet coefficient block
will comprise a 2x4 coefficient quadrant. In processing the sub-quadrants of
such a block, the sub-quadrants may be arbitrarily designated to emphasize the

5   most important information. For example, in one embodiment of the invention
utilizing a non-truncated 4x8 block, the LLLL sub-quadrant comprises wavelet
coefficient X_0, the LLLH sub-quadrant comprises wavelet coefficients X_4 and X_5, ,
the LLHL sub-quadrant comprises wavelet coefficients X_1 and X_2, and the LLHH
sub-quadrant comprises wavelet coefficients X_3, X_6 and X_7. Alternatively, the

10 generated wavelet coefficients for each sub-quadrant may be reduced to a single
coefficient by averaging or otherwise combining the generated wavelets.

FIG. 6 depicts a block diagram of a compression process according to the
invention. Specifically, FIG. 6 depicts a block diagram of a process 600 for
decomposing pixel domain information into wavelet domain information.

15 Specifically, four pixels 610-613 are processed according to a standard Haar
transform to produce four Haar wavelet transform coefficients 630-633 at a first
level of decomposition. Additionally, the first two wavelet coefficients 630, 631 a
second level decomposition utilizing the Haar wavelet transform is used to
process to produce a pair of DCT domain coefficients.

20 A first level decomposition utilizing a Haar wavelet transform is
performed on the pixel block to produce a corresponding Haar wavelet coefficient
block. Specifically, pixel 610 is added to pixel 611 by a first adder 620 to produce
a first wavelet coefficient 630 (i.e., X_0). Pixel 612 is added to pixel 613 by a
second adder 622 to produce a second wavelet coefficient 631 (i.e., X_1). Pixel 611

25 is subtracted from pixel 610 to produce third wavelet coefficient 632 (i.e., X_2).
Pixel 613 is subtracted from pixel 612 by a second subtractor 623 to produce a
fourth wavelet coefficient 633 (i.e., X_3). Thus, the pixel block (x_0-x_3) is subjected
to a first level of decomposition to produce a corresponding Haar wavelet
coefficient block (X_0-X_3).

30 Additionally, a second level decomposition is performed on the first and
second wavelet coefficients 630 and 631. Specifically, wavelet coefficient 630 is
added to wavelet coefficient 631 by a third adder 641 to produce a fifth wavelet

coefficient 650 (i.e., Y_0); and wavelet coefficient 631 is subtracted from wavelet coefficient 630 by a third subtractor 642 to produce a sixth wavelet coefficient 651 (i.e., Y_1). Thus, pixel 610 to 613 are processed using the Haar transform to produce first level decomposition Haar wavelet coefficients 632 and 633, and
5   second level decomposition wavelet coefficient 650 and 651. By processing pixel 610 through 613 in this matter, the resulting wavelet coefficients 632, 633, 650 and 651 are produced which represent the same pixel information in the wavelet domain.

The above-described process 600 is suitable for use in, e.g., the
10   compression module 200 of FIG. 1 or FIG. 2 in non-DCT compression mode. Similarly, the decompression module 300 of FIG. 1 or FIG. 3 in non-DCT compression mode will perform a mirror image of the above-described process 600 to extract, from the various coefficients, the original pixel block compressed by compression module 200.

15   FIG. 7 depicts a block diagram of a compression process according to the invention. Specifically, FIG. 7 depicts a block diagram of a process 700 for decomposing pixel domain information into wavelet domain information. A pixel block comprising pixels 710 through 717 (i.e., x_0-x_7) is subjected to a DCT transform process 720 to produce a corresponding DCT coefficient block
20   comprising DCT coefficients 730 through 737 (i.e., X_0-X_7). The high order DCT coefficients 734 through 737 are truncated, while the low order DCT coefficients 730 through 733 are coupled to respective scaling units 740 through 743. Scaling units 740 through 743 scale, respectively, low order DCT coefficients 730 through 733 by a factor of 1 divided by the square root of 2 (i.e., 0.7071).

25   The scaled coefficients (i.e., the coefficients at the output of a respective scaling unit) are then subjected to a modified Haar transform (i.e., a Haar transform in which the order of additions and subtractions have been modified from that of a standard Haar transform). Specifically, the coefficient produced by scaler 740 is added to the coefficient produced by scaler 741 by a first adder
30   750 to produce a first wavelet coefficient 760 (i.e., Z_0). The coefficient produced by scaler 741 is subtracted from the coefficient produced by scaler 740 by a first subtractor 751 to produce a second wavelet coefficient 761 (i.e., Z_1). The

coefficient produced by scaler 742 is added to the coefficient produced by scaler 743 by a second adder 752 to produce a third wavelet coefficient 762 (i.e., Z_2). The coefficient produced by scaler 743 is subtracted from the coefficient produced by scaler 742 by a second adder 753 to produce a fourth wavelet coefficient 763

5   (i.e., Z_3).

A second level wavelet decomposition is performed on the first two wavelet coefficients 760 and 761 as follows. The first wavelet coefficient 760 is added to the second wavelet coefficient 761 by an adder 770 to produce a first second-level decomposition wavelet coefficient 780. The second wavelet coefficient 761 is

10   subtracted from the first wavelet coefficient 760 by a subtractor 771 to produce a second-level decomposition wavelet coefficient 781.

Thus, the pixel block represented by pixel 710 through 717 has been compressed into two first-level decomposition wavelet coefficients (i.e., 762 and 763) and two second-level decomposition wavelet coefficients (780 and 781).

15   The above-described process 700 is suitable for use in, e.g., the compression module 200 of FIG. 1 or FIG. 2 in DCT compression mode. Similarly, the decompression module 300 of FIG. 1 or FIG. 3 in DCT compression mode will perform a mirror image of the above-described process 700 to extract, from the various coefficients, the original pixel block compressed by compression

20   module 200.

The present invention can be embodied in the form of computer-implemented processes and apparatuses for practicing those processes. The present invention also can be embodied in the form of computer program code embodied in tangible media, such as floppy diskettes, CD-ROMs, hard

25   drives, or any other computer readable storage medium, wherein , when the computer program code is loaded into and executed by a computer, the computer becomes an apparatus for practicing the invention. The present invention can also be embodied in the form of computer program code, for example whether stored in a storage medium, loaded into and/or executed by a computer, or

30   transmitted over some transmission medium, such as over electrical wiring or cabling, through fiber optics, or via electromagnetic radiation, wherein, when the computer program code is loaded into and executed by a computer, the computer

19

becomes an apparatus for practicing the invention. When implemented on a general-purpose microprocessor, the computer program code segments configure the microprocessor to create specific logic circuits.

Although various embodiments which incorporate the teachings of the present invention have been shown and described in detail herein, those skilled in the art can readily devise many other varied embodiments that still incorporate these teachings.

What is claimed is:

1.   In a system for processing an MPEG-like video stream, a method comprising the steps of:

5   transforming, using a Haar wavelet transform, a pixel block to form a Haar coefficient block;

quantizing, using a plurality of scaling factors, said Haar coefficient block to form a respective quantized Haar coefficient block, said plurality of scaling factors being selected to preferentially allocate an available bit budget to

10 quantized Haar coefficients representing relatively low vertical and horizontal spatial frequency components of said pixel block;

packing each of said quantized Haar coefficients and at least one of said plurality of scaling factors into a respective word; and

storing said word in a memory module within said decoding system.

15

2.   The method of claim 1, wherein said steps of transforming, quantizing, packing and storing are repeated for each pixel block forming an image frame, such that said memory module stores an entire compressed image frame.

20 3.   The method of claim 1, further comprising the steps of:

reading, from said memory module, said stored word;

unpacking said stored word to retrieve said quantized Haar coefficients and said at least one of said plurality of scaling factors;

inverse quantizing, using said retrieved at least one of said plurality of

25 scaling factors, said Haar coefficient block to form a respective inverse quantized Haar coefficient block;

inverse transforming, using said Haar wavelet transform, said inverse quantized Haar coefficient block to form a respective pixel block.

30 4.   The method of claim 3, wherein said steps of transforming, quantizing, packing and storing are repeated for each pixel block forming an image frame, such that said memory module stores an entire compressed image frame.

21

5.   The method of claim 1, wherein said step of transforming comprises the steps of:

        transforming, using a discrete cosine transform (DCT) said pixel block to

5   form a DCT coefficient block; and

        transforming, using said modified Haar wavelet transform, said DCT coefficient block to form said Haar coefficient block.

6.   The method of claim 1, wherein said step of transforming comprises the

10 steps of:

        transforming, using a discrete cosine transform (DCT) said pixel block to form an DCT coefficient block;

        truncating a plurality of high order DCT coefficients within said DCT coefficient block to form a truncated DCT coefficient block; and

15   transforming, using said modified Haar wavelet transform, said truncated DCT coefficient block to form said Haar coefficient block.

7.   The method of claim 6, wherein said step of truncation reduces the number of coefficients in said DCT coefficient block by a factor of at least two.

20

8.   The method of claim 1, wherein said step of quantization comprises the steps of:

        associating said Haar wavelet coefficients with one of a low vertical and low horizontal spatial frequency quadrant (LL), a low vertical and high

25 horizontal spatial frequency quadrant (LH), a high vertical and low horizontal spatial frequency quadrant (HL), and a high vertical and high horizontal spatial frequency quadrant (HH), said LL quadrant comprising a DC sub-quadrant and a plurality of non-DC sub-quadrants;

        quantizing, using a first bit allocation, said Haar coefficients associated

30 with said DC sub-quadrant;

        quantizing, using a second bit allocation, said Haar coefficients associated with said non-DC sub-quadrants;

quantizing, using a third bit allocation, said Haar coefficients associated with said HL and HH quadrants;

quantizing, using a fourth bit allocation, said Haar coefficients associated with said LH quadrant, where said fourth bit allocation is increased if more than
5   a predetermined number of LL, HL and HH quadrant coefficients are quantized to zero.

9.   The method of claim 8, wherein, in the case of said pixel blocks comprising a 4x4 pixel block:
10   said first bit allocation comprises an eight bit allocation;

said second bit allocation comprises a four bit allocation and a sign bit allocation;

said third bit allocation comprises a two bit allocation and a sign bit allocation; and
15   said fourth bit allocation comprises either a two bit allocation and a sign bit allocation, or a three bit allocation and a sign bit allocation.

10.   The method of claim 8, wherein:

said Haar coefficients associated with said DC sub-quadrant are
20   quantized using a first scaling factor;

said Haar coefficients associated with said non-DC sub-quadrants are quantized using a second scaling factor;

said Haar coefficients associated with said HL and HH quadrants are quantized using a third scaling factor; and
25   said Haar coefficients associated with said LH quadrant are quantized using one of said third scaling factor or a fourth scaling factor.

11.   The method of claim 10, wherein:

said first scaling factor is a predetermined value;
30   said second scaling factor comprises said predetermined value multiplied by 2^m;

said third scaling factor comprises said predetermined value multiplied by 2^n; and

said fourth scaling factor comprises said predetermined value multiplied by $2^{(n)}$ or $2^{(n-1)}$.

5

12.   The method of claim 8, wherein:

said Haar coefficients associated with said DC sub-quadrant are quantized using a scaling factor of four;

said Haar coefficients associated with said non-DC sub-quadrants are

10   quantized using a scaling factor of $4*2^n$;

said Haar coefficients associated with said HL and HH quadrants are quantized using a scaling factor of $4*2^n$; and

said Haar coefficients associated with said LH quadrant are quantized using a scaling factor of $4*2^n$ or $4*2^{(n-1)}$.

15

13.   The method of claim 12, wherein:

said steps of transforming, quantizing, packing and storing are repeated for each of a plurality of pixel blocks; and

said scaling factor variables n and m are recalculated for each respective

20   Haar coefficient block.

14.   In a system for processing one or more images within a video information stream, where each image comprises a plurality of pixel blocks, a method for processing each of said pixel blocks comprising the steps of:
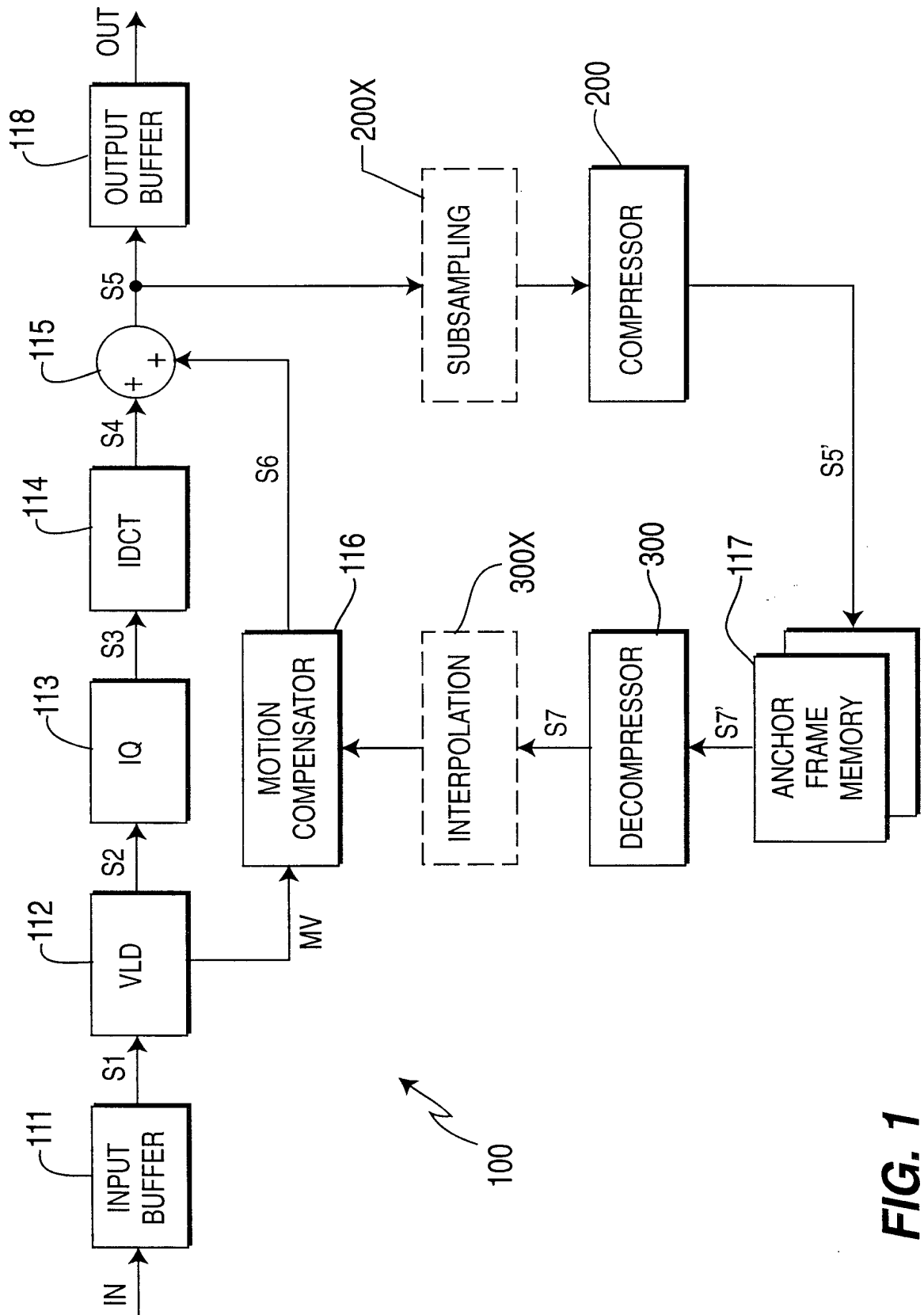
25   processing, according to a Haar wavelet transform, said pixel blocks to produce a respective Haar wavelet coefficient blocks;

associating said Haar wavelet coefficients within said with one a low vertical and low horizontal spatial frequency quadrant (LL), a low vertical and high horizontal spatial frequency quadrant (LH), a high vertical and low

30   horizontal spatial frequency quadrant (HL), and a high vertical and high horizontal spatial frequency quadrant (HH), said LL quadrant comprising a DC sub-quadrant and a plurality of non-DC sub-quadrants;
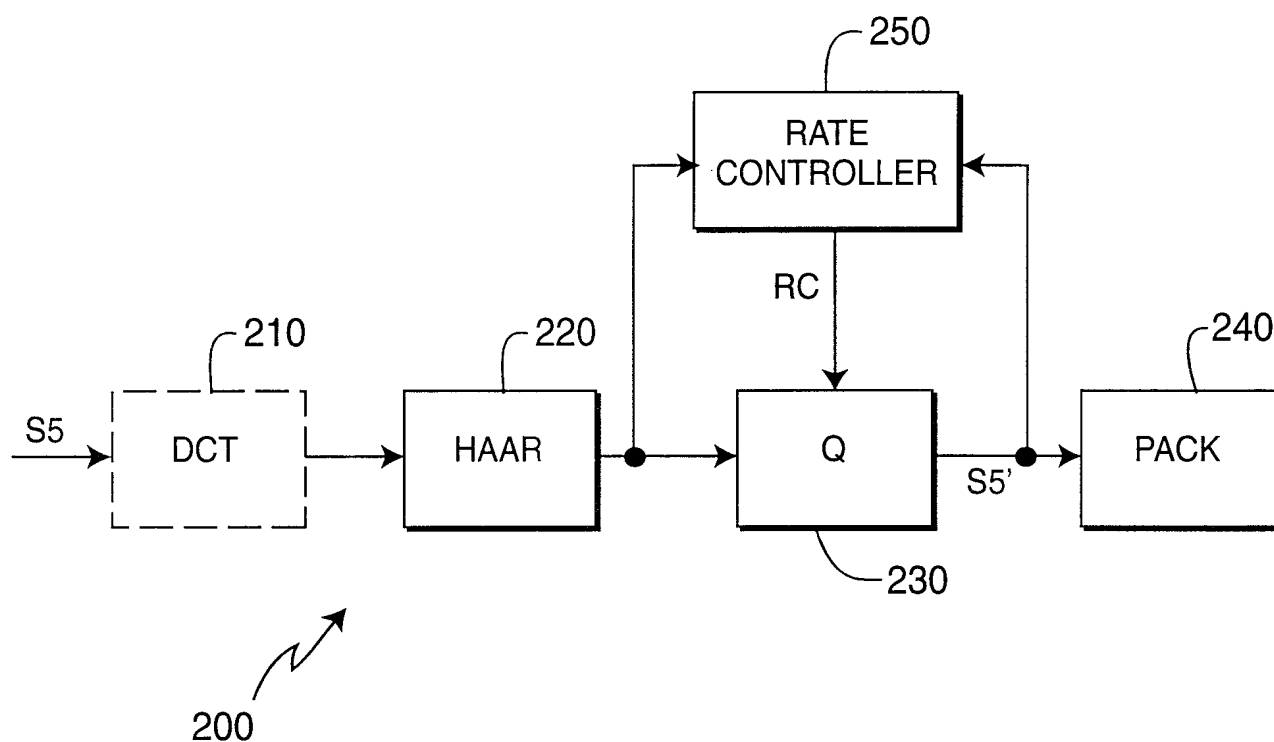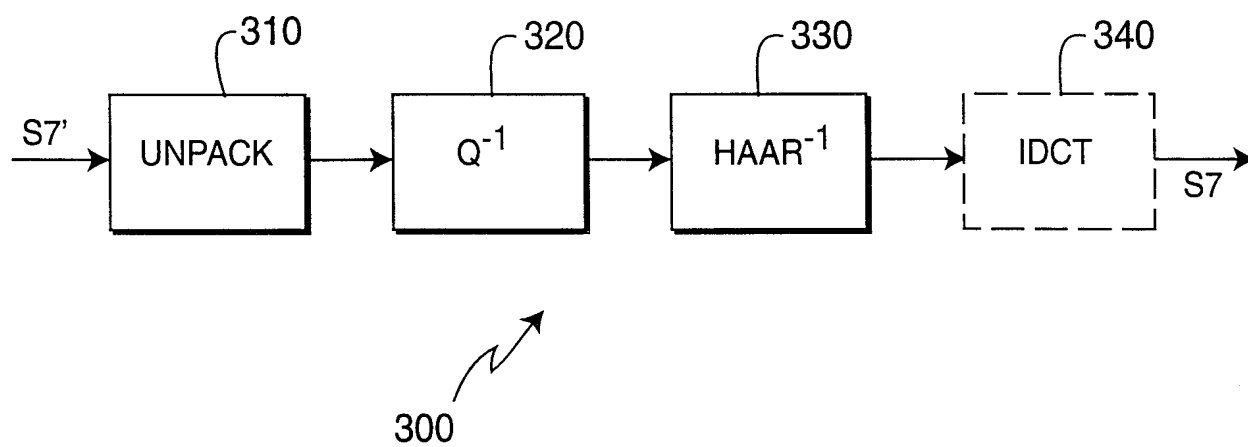
preferentially quantizing, in the following order, said Haar coefficients associated with said DC sub-quadrant, said Haar coefficients associated with said non-DC sub-quadrants, said Haar coefficients associated with said LH quadrant, and said Haar coefficients associated with said HL and HH

5   quadrants, where said Haar coefficients associated with said DC sub-quadrant are provided a relatively large bit allocation, and said LH quadrant coefficients are quantized according to a bit allocation that is increased if more than a predetermined number of LL, HL and HH quadrant coefficients are quantized to zero; and

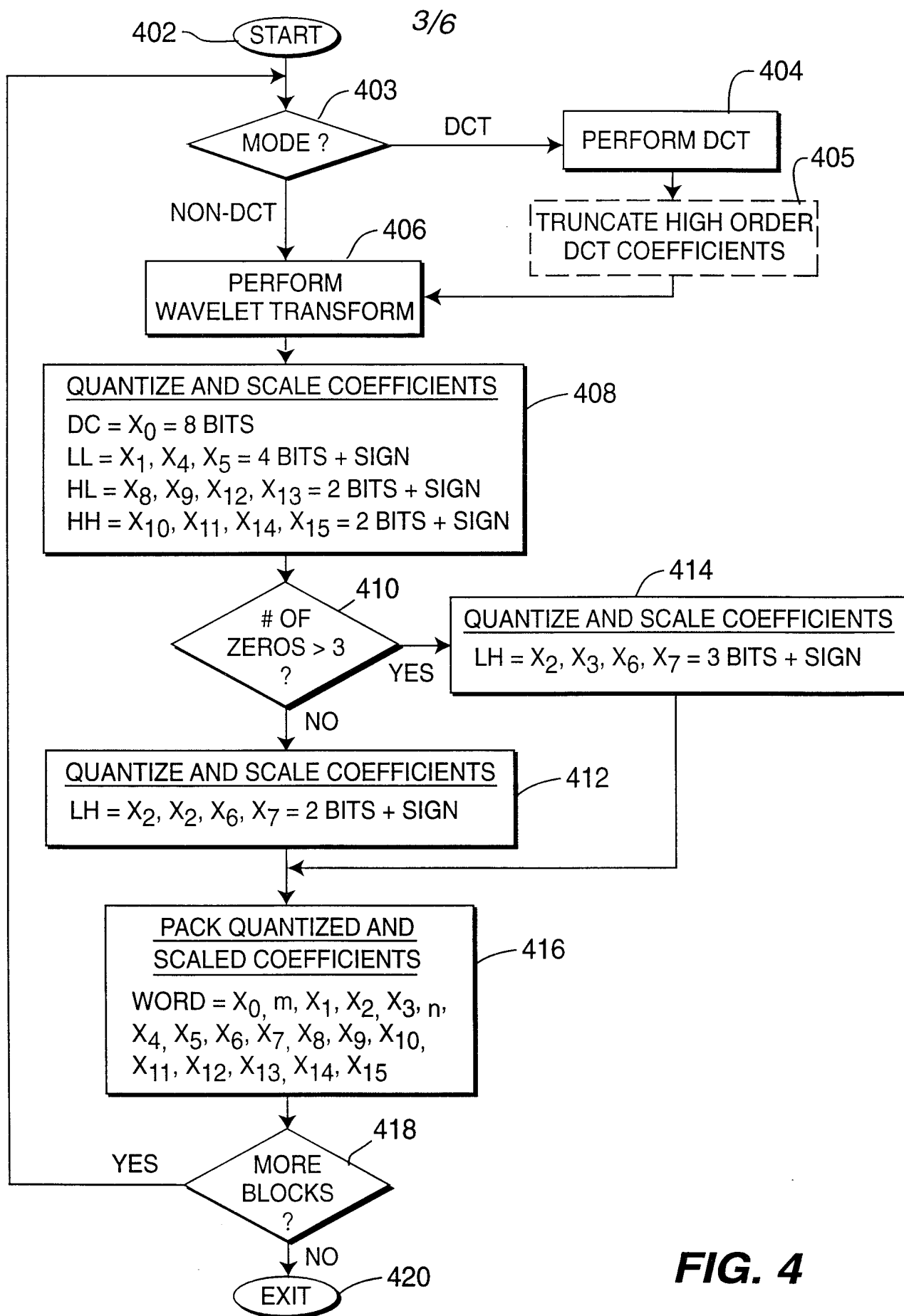10   packing said quantized Haar coefficients and indicia of associated quantization scaling factors into a word.
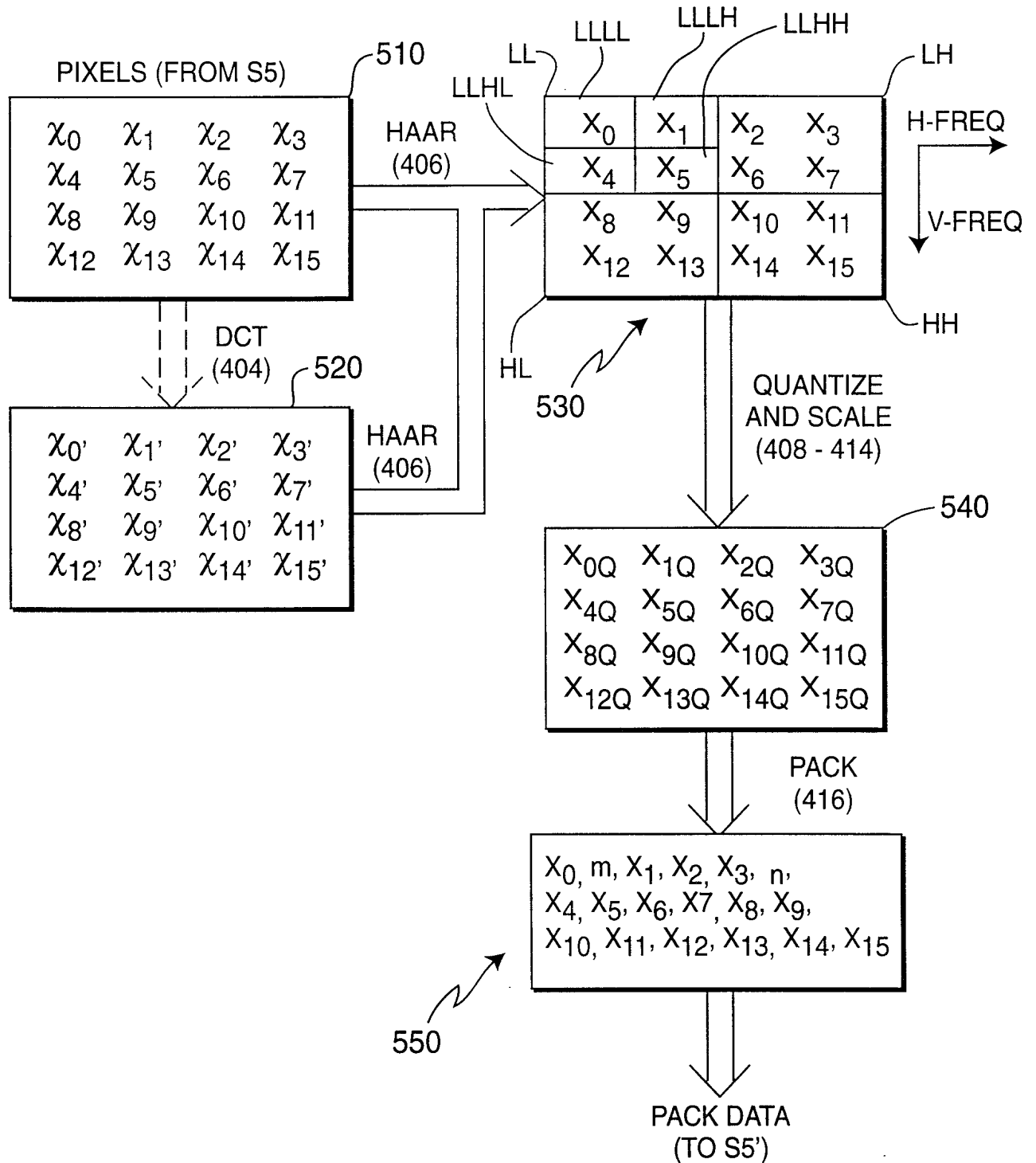
1/6

FIG. 1
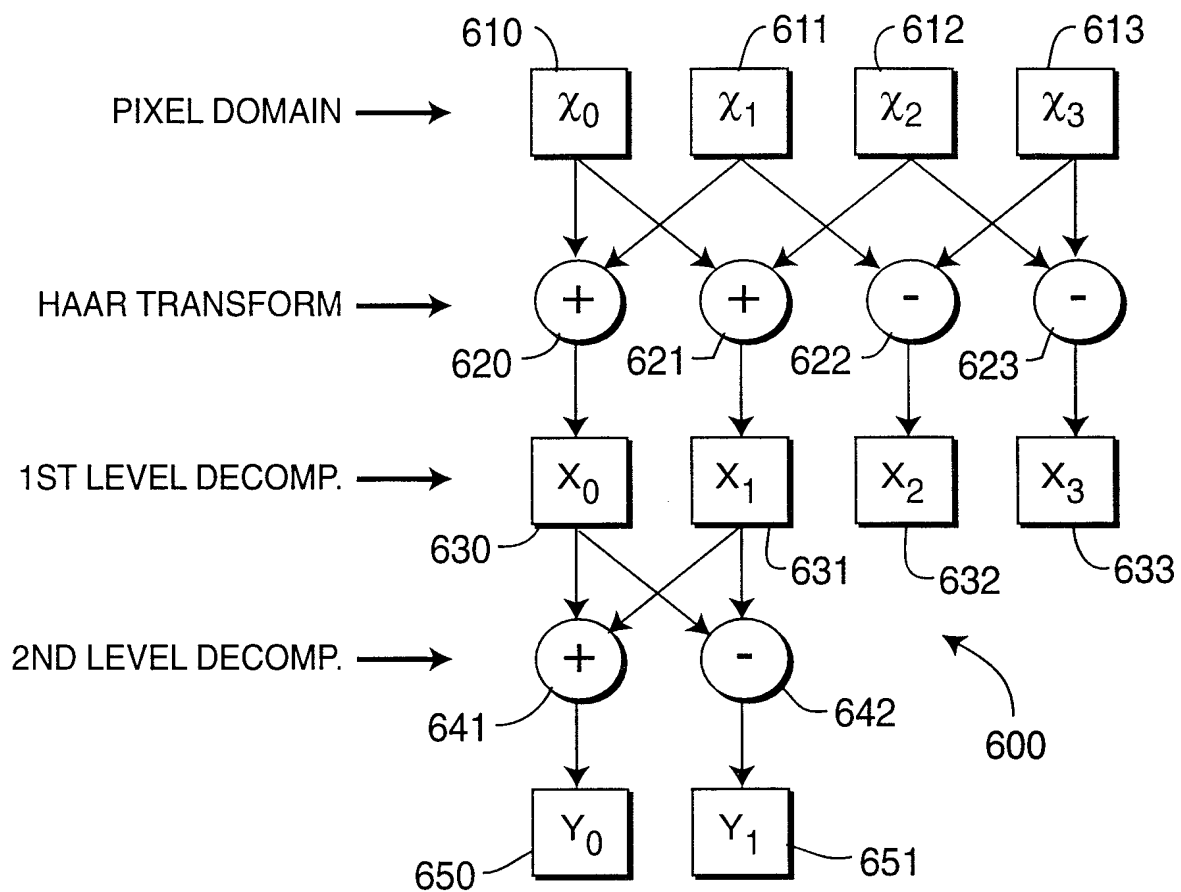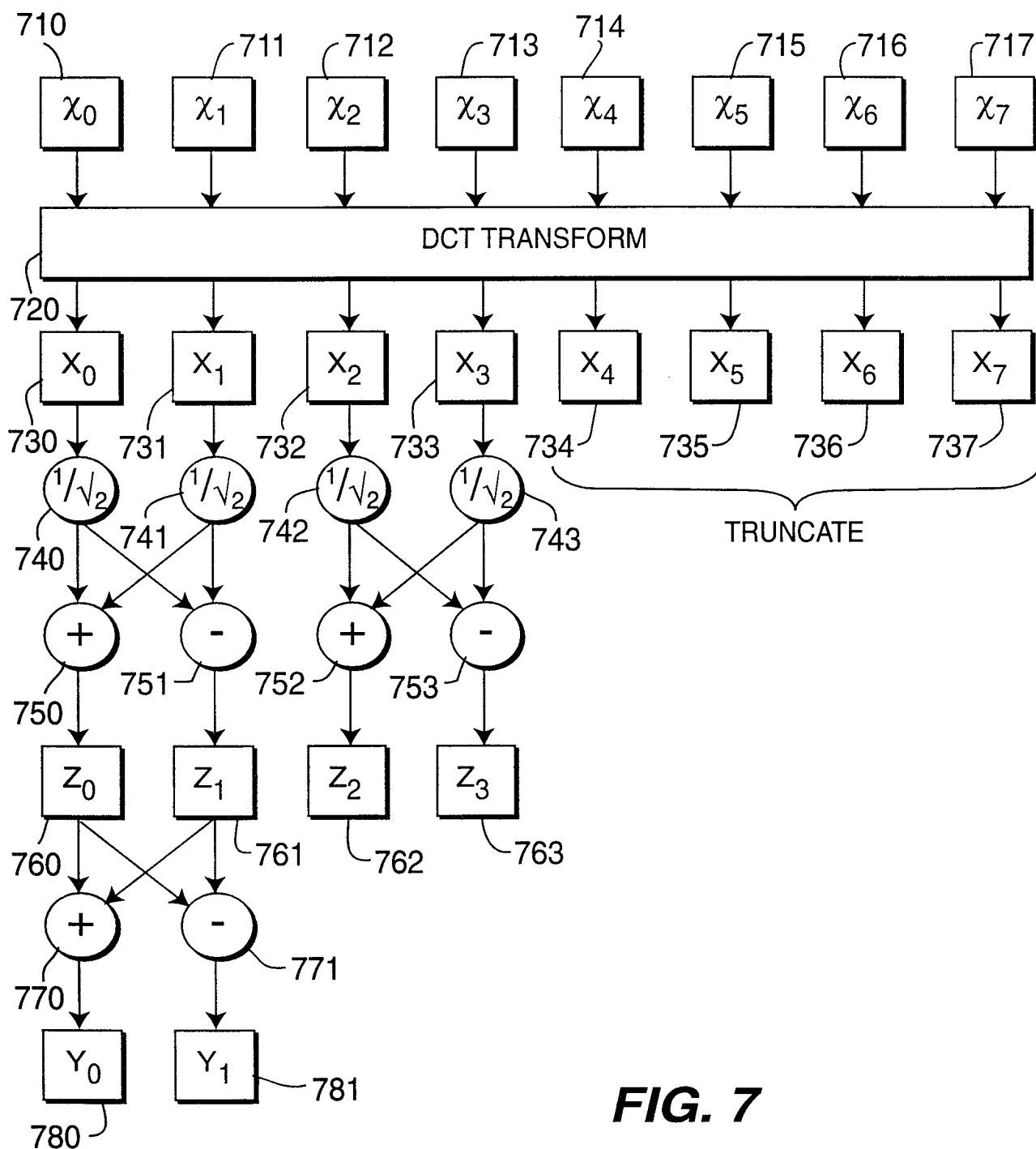
FIG. 2

FIG. 3

START — 402

403

MODE ? — DCT → PERFORM DCT — 404

405 — TRUNCATE HIGH ORDER DCT COEFFICIENTS

NON-DCT ↓

406 — PERFORM WAVELET TRANSFORM

QUANTIZE AND SCALE COEFFICIENTS — 408

$DC = X_0 = 8$ BITS

$LL = X_1, X_4, X_5 = 4$ BITS + SIGN

$HL = X_8, X_9, X_{12}, X_{13} = 2$ BITS + SIGN

$HH = X_{10}, X_{11}, X_{14}, X_{15} = 2$ BITS + SIGN

410 — # OF ZEROS > 3 ?

414 — QUANTIZE AND SCALE COEFFICIENTS

$LH = X_2, X_3, X_6, X_7 = 3$ BITS + SIGN

YES →

NO ↓

QUANTIZE AND SCALE COEFFICIENTS — 412

$LH = X_2, X_2, X_6, X_7 = 2$ BITS + SIGN

PACK QUANTIZED AND SCALED COEFFICIENTS — 416

$WORD = X_0, m, X_1, X_2, X_3, n,$
$X_4, X_5, X_6, X_7, X_8, X_9, X_{10},$
$X_{11}, X_{12}, X_{13}, X_{14}, X_{15}$

418 — MORE BLOCKS ?

YES ↑

NO ↓

EXIT — 420

FIG. 4

FIG. 5

FIG. 6

FIG. 7

INTERNATIONAL SEARCH REPORT

A. CLASSIFICATION OF SUBJECT MATTER
IPC 6   H04N7/26   H04N7/50

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC 6   H04N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication. where appropriate, of the relevant passages	Relevant to claim No.
X	US 5 434 567 A (MACK WALTER ET AL) 18 July 1995 (1995-07-18) column 4, line 51 – column 5, line 6 column 7, line 14 – column 8, line 21 column 10, line 38 – column 10, line 41 column 11, line 39 – column 12, line 60	1,2,5,6
A		3,4,7-14

X	EP 0 794 673 A (MITSUBISHI ELECTRIC CORP) 10 September 1997 (1997-09-10) column 3, line 27 – column 3, line 59 column 6, line 53 – column 8, line 17 claims 1,7	1-4
A		5-14

	-/--	

| X | Further documents are listed in the continuation of box C. |
| X | Patent family members are listed in annex. |

° Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
6 August 1999	30/08/1999

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl. Fax: (+31-70) 340-3016	Fassnacht, C

Form PCT/ISA/210 (second sheet) (July 1992)

1

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category°	Citation of document, with indication,where appropriate, of the relevant passages	Relevant to claim No.
A	CLARKE R J: "TRANSFORM CODING OF IMAGES" 1985 , 2ND PRINTING 1990, ACADEMIC PRESS, "MICROELECTRONICS AND SIGNAL PROCESSING" SERIES , LONDON (GB), SECTIONS 3.5.4 , 7.4 XP002111367 * page 305, eq.(7.24) *	1-14
A	US 5 706 220 A (RAMACHANDRAN LOGANATH ET AL) 6 January 1998 (1998-01-06) column 3, line 35 - column 4, line 46	1-4,8-14
A	ALBANESI M G ET AL: "IMAGE COMPRESSION BY THE WAVELET DECOMPOSITION" EUROPEAN TRANSACTIONS ON TELECOMMUNICATIONS AND RELATED TECHNOLOGIES, vol. 3, no. 3, 1 May 1992 (1992-05-01), pages 265-274, XP000304925 ISSN: 1120-3862 * sections 1, 2 *	1-4,8-14

1

INTERNATIONAL SEARCH REPORT

Information on patent family members

Patent document cited in search report		Publication date	Patent family member(s)		Publication date
US 5434567	A	18-07-1995	NONE		
EP 0794673	A	10-09-1997	CA	2185753 A	18-03-1998
			AU	693718 B	02-07-1998
			AU	6568896 A	11-09-1997
			CN	1166110 A	26-11-1997
			FR	2745679 A	05-09-1997
			JP	10098731 A	14-04-1998
			US	5701158 A	23-12-1998
US 5706220	A	06-01-1998	NONE		